AL/HR-TP-1996-0003

ARMSTRONG LABORATORY

# OPTIMAL PERSONNEL ASSIGNMENT: AN APPLICATION TO AIR FORCE PILOTS


Frederick M. Siem
William E. Alley

HUMAN RESOURCES DIRECTORATE
MANPOWER AND PERSONNEL RESEARCH DIVISION
7909 Lindbergh Drive
Brooks AFB TX 78235-5352





March 1996

Interim Technical Paper for Period February 1993 - December 1993




AIR FORCE MATERIEL COMMAND
BROOKS AIR FORCE BASE, TEXAS

# NOTICES

Publication of this paper does not constitute approval or disapproval of the ideas or findings. It is published in the interest of scientific and technical information (STINFO) exchange.

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any other way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell patented invention that may in any way be related thereto.

The Office of Public Affairs has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

Frederick M Siem

FREDERICK M. SIEM
Personnel Research Psychologist
Aircrew Selection Research Branch

PATRICK C. KYLLONEN
Technical Director
Manpower & Personnel Research Division

GARY D. ZANK, Colonel, USAF
Chief, Manpower & Personnel Research Division

Please notify this office, AL/HRPP, 7909 Lindbergh Drive, Brooks AFB TX 78235-5352, if your address changes, or if you no longer want to receive our techinical reports. You may write or call the STINFO office at DSN 240-3853 or commercial (210) 536-3853.

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. AGENCY USE ONLY (Leave blank) | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | March 1996 | Interim Paper - October 1993-December 1994 |

**4. TITLE AND SUBTITLE**

Optimal Personnel Assignment: An Application to Air Force Pilots

**5. FUNDING NUMBERS**

PE - 62205F
PR - 7719, 1123
TA - 25, A1
WU - 01,01

**6. AUTHOR(S)**

Siem, Frederick M.
Alley, William E.

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Armstrong Laboratory
Human Resources Directorate
Manpower and Personnel Research Division
7909 Lindbergh Drive
Brooks AFB TX 78235-5352

**8. PERFORMING ORGANIZATION REPORT NUMBER**

AL/HR-TP-1996-0003

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**

Armstong Laboratory Technical Monitor: Dr. Siem (210) 536-3956
Also published in Military Psychology, 7(4), 253-263

**12a. DISTRIBUTION/AVAILABILITY STATEMENT**

Approved for public release; distribution is unlimited.

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** *(Maximum 200 words)*

A study was conducted to examine the potential utility of optimally assigning Air Force pilots to training tracks without benefit of actual training outcomes. The resulting assignment solution indicated that a) there was sufficient agreement among pilots to form coherent selection policies that differed across types of aircraft, and b) mean predicted performance could be improved about one-third standard deviation relative to random allocation. Follow-up research is discussed.

**14. SUBJECT TERMS**

Prdictive performance
Selection policies
Training tracks

**15. NUMBER OF PAGES**

17

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | UL |

## CONTENTS

## TABLES

**Table No.**



## FIGURES

**Figure No.**

# PREFACE

This research was completed under Work Unit 77191845 in support of a Request for Personnel Research (RPR 80-02, Selection for Flying Training Tracks) submitted by Air Force training program managers. This paper is intended to serve as interim documentation regarding the use of optimal assignment algorithms to improve pilot track assignment.

MILITARY PSYCHOLOGY, 7(4), 253–263

# Optimal Personnel Assignment: An Application to Air Force Pilots

Frederick M. Siem and William E. Alley
*Armstrong Laboratory Human Resources Directorate*
*Brooks Air Force Base, Texas*

A study was conducted to examine the potential utility of optimally assigning Air Force pilots to training tracks without benefit of actual training outcomes. The resulting assignment solution indicated that (a) there was sufficient agreement among pilots to form coherent selection policies that differed across types of aircraft and (b) mean predicted performance could be improved about one third of a standard deviation relative to random allocation. Follow-up research is discussed.

The military has a long history of employing personnel classification techniques to improve initial assignment decisions. In 1942, the Army Air Forces designed a system for allocating military applicants to pilot, navigator, and bombardier training based on scores from a multiple aptitude battery (Flanagan, 1948). Although the problem could be clearly specified at the time, only approximate solutions were available for optimizing the process (Thorndike, 1947). It was not until the late 1940s and early 1950s that psychometric advances in the field, exemplified by the work of Brogden (1954), Horst (1956), and Ward (1958), could be coupled with developments in operation research (i.e., linear programming) so that definitive solutions could be obtained. The more recent history of linear programming algorithms for personnel classification are discussed in Johnson and Zeidner (1990).

In the original World War II context, as with most applications, personnel are initially assigned to training programs without benefit of knowledge about how test scores relate to training outcomes. Trainees are followed over a period of time, and when sufficient criterion data are assembled, empirical prediction systems can be generated within each program to serve as the

Requests for reprints should be sent to Frederick M. Siem, AL/HRMA, 7909 Lindbergh Drive, Brooks Air Force Base, TX 78235–5352.

basis for establishing classification guidelines. A different problem arose in connection with early planning for a recent pilot training initiative in which specialized primary flight instruction was to replace a common program for each of four categories of pilot trainee: fighter, bomber, tanker, or transport. Training managers wanted to develop a classification procedure that could be used in the interim. Based on previous research, policy capturing and judgment analysis (Bottenberg & Christal, 1968; Christal, 1968; Naylor & Wherry, 1965) provided an approach for creating synthetic prediction equations that would serve as interim criteria until the program had matured sufficiently to employ empirical equations. At issue was (a) whether experienced pilots could differentiate from among applicants those who would be best suited for assignment to specific training tracks and (b) whether and to what extent expected performance gains were possible by employing these data in an optimal classification process. *Optimal assignment* in this context refers to maximizing the mean predicted performance for a group of job candidates assigned to different job categories. That is, we want a rule, or objective function, by which to match job candidates to job categories that makes the most utility of the human resources available. Maximizing mean predicted performance across job categories is just one rule by which to make personnel assignments. Other rules might be to maximize performance in one job or to randomly assign individuals to jobs. Johnson and Zeidner (1990) provided a detailed discussion of the use and nature of various classification algorithms.

Ward (1958) provided a simplified example of the job assignment problem addressed by multiple attribute theory (see Figure 1). One rule for job assignment would be to enter each person sequentially into the job for which he or she is most qualified. Thus, Person A would be assigned to Job 1, Person B would be assigned to Job 2, and finally Person C would be assigned to Job 3. The result would be a mean predicted-performance score of $(9 + 2 + 2)/3 = 4.33$. Another strategy would be to consider all three applicants at the same time, but to consider the jobs one at a time, so that each job was assigned in turn to the most qualified applicant. This strategy, considering the jobs in numeric order, would result in a predicted-performance score of $(9 + 6 + 1)/3 = 5.33$, some improvement over the first strategy.

Both of the aforementioned strategies can be considered single attribute rules. In one case, only persons are considered; in the second, only jobs. Now

| | *Payoff Values for Alternative Jobs* | | |
|---|---|---|---|
| *Person* | *Job 1* | *Job 2* | *Job 3* |
| A | 9 | 8 | 7 |
| B | 6 | 2 | 1 |
| C | 7 | 6 | 2 |

FIGURE 1 Example of multiple attribute assignment problem.

consider an alternate assignment scheme that simultaneously considers both persons and jobs in order to generate the maximum predicted performance. Such a strategy can be considered a multiple attribute strategy. With the multiple attribute strategy, Person A would be assigned to Job 3, Person B to Job 1, and Person C to Job 2. The mean predicted score by this rule is (7 + 6 + 6)/3 = 6.33.

Although procedures for maximizing mean predicted performance across job categories have been available for some time (Johnson & Zeidner, 1990), potential applications are somewhat limited by the situation required to use such data, namely one in which a group of candidates are simultaneously assigned to different jobs. Such an approach has been given limited implementation (Johnson & Zeidner, 1990) and, as a consequence, little is known about the utility that may exist in practice for various applications of the procedure. The purpose of the present study was to examine the utility of optimal classification procedures for assignment of Air Force pilot candidates to four separate training tracks prior to the availability of actual training outcomes.

## METHOD

### Participants

The participants in the study were 57 male Air Force Instructor Pilots (IPs) who served as Subject Matter Experts (SMEs). Thirteen of the SMEs were fighter IPs; 11 SMEs were bomber IPs; 16 SMEs were tanker IPs; and the remaining 17 SMEs were transport IPs. The SMEs typically had several thousand hours of experience piloting jet aircraft (range: 2,000–10,000 hr).

### Measures

The main criterion measure of interest was predicted training performance in four different types of aircraft: bomber, fighter, tanker, and transport. To develop predicted-performance measures for each aircraft type, a policy-capturing exercise was conducted (Christal, 1968; Naylor & Wherry, 1965). The stimulus materials presented to the SMEs consisted of data cards containing information about 200 pilot candidates on several dimensions (see Table 1).

The data cards included information about four aptitude measures from the Air Force Officer Qualifying Test (AFOQT; Skinner & Ree, 1987), a paper-and-pencil aptitude test used for Air Force pilot candidate selection since 1955. The AFOQT consists of 16 subtests that for operational purposes are combined into five composites. The scores used in the present study were the Pilot, Navigator–Technical, Verbal, and Quantitative composites. The

TABLE 1
Variables Used in Policy-Capturing Exercise

| *Variable* | *Construct Measured* |
|---|---|
| Information-processing speed | Ability to respond quickly to information |
| Information-processing accuracy | Ability to respond accurately to information |
| Resource allocation | Ability to perform two tasks at same time |
| Hand–eye coordination | Stick-and-rudder skills |
| Mental flexibility | Open-mindedness |
| Tolerance for monotony | Ability to perform routine tasks for extended period |
| Leadership | Interpersonal and communication skills |
| Timing | Ability to estimate rate of movement |
| Procedural memory | Ability to remember and apply complex rules |
| Mental visualization | Ability to compare complex visual figures |
| Grade point average | College GPA on a 4.0 scale |
| AFOQT pilot | Aptitude for completion of pilot training |
| AFOQT navigator–technical | Aptitude for completion of navigator training |
| AFOQT verbal | Reading comprehension, word relationships |
| AFOQT quantitative | Understanding of math relationships |
| PPL | Private Pilot License |
| Technical degree | College degree in engineering, natural sciences, or computer sciences |
| Aircraft preference | Preference to fly either bomber, fighter, tanker, or transport aircraft |

*Note.* AFOQT = Air Force Officer Qualifying Test.

fifth composite, Academic Aptitude, was not used in the present study because of space limitations on the stimulus materials and because it is derived from two other composites, Verbal and Quantitative, rendering the information redundant.

The data cards also contained 10 scores from the Basic Attributes Tests (BAT; Carretta, 1990), a computer-administered battery of psychomotor, cognitive, and personality tests. Five of the scores were composites based on seven tests that have been experimentally validated against pilot training performance for samples of Air Force pilot candidates. The five composites were (a) information-processing speed, based on response latency scores from three BAT tests (Item Recognition, Mental Rotation, and Encoding Speed); (b) information-processing accuracy, based on the same three tests; (c) resource allocation, based on measures from the BAT Time Sharing test; (d) hand–eye coordination, based on two BAT psychomotor tests (Two-Hand Coordination and Complex Coordination); and (e) mental flexibility, based on scores from the Self-Crediting Word Knowledge test.

The AFOQT and BAT scores previously described were generated from archival data on student pilots tested on both the AFOQT and the BAT. Because data for five of the BAT tests were not available for participants in the archival database, scores for the following constructs were generated synthetically: Tolerance for Monotony, Leadership, Timing, Procedural Memory, and Mental Visualization. A rectangular distribution of scores was created, and decile scores were randomly assigned to the 200 records.

Both the AFOQT and BAT scores were represented on a 10-point scale representing single-digit percentile or decile scores (1%–10% = 1, etc.). The AFOQT scores were labeled with the acronym for that test, because pretesting demonstrated adequate familiarity with the test (most pilots had been selected based on scores from the AFOQT). Because pretesting also demonstrated a relative lack of familiarity with the BAT battery, the scores representing the BAT were labeled with names of the constructs measured by the scores—that is, Hand–eye Coordination, Leadership, and so forth.

Finally, the data cards included demographic variables: possession of a civilian pilot license, technical major in college, college grade point average, and aircraft assignment preference. For analytical purposes, the preference measure was converted to four binary variables, each one representing assignment preference for one of four types of aircraft.

## Procedure

An experimenter explained the purpose of the card-sorting exercise to the participant SMEs. The nature of the tests used to generate the scores on the applicant profile cards were explained in detail. The SMEs were then given information on the 200 applicants. Each SME was asked to rank order the candidates in terms of expected performance in the SME's particular aircraft type. The 200 cards were divided into four groups of 50 to minimize the burden of the ranking task. Thus, each SME rank ordered the candidates one time only, and the rank order (from 1 to 50) served as the performance criterion measure. For analyses, the rank orders were recoded so that 50 was the highest score and 1 the lowest.

## Analysis

The first stage of analysis examined the rankings by aircraft type for interrater reliability using software developed for occupational task inventory ratings (Christal & Weismuller, 1976). The next stage of analysis was designed to address the issue of whether each of the SMEs was internally consistent in his policy for rank ordering the candidates. Intrarater consistency analyses involved development of separate regression equations for each SME, with the ranking criterion regressed on the variables included in the data cards. Following conventional practice, a high multiple correlation between each rater's ranking and the set of predictor variables served as an index of internal consistency (Dougherty, Ebert, & Callender, 1986). That is, if an SME failed to use a consistent policy, then one would expect to find no relation between the scores on the applicant profiles and an individual SME's rankings. Next the regression equations, or policies used by each

rater, were used in a hierarchical cluster analysis to determine the number of different policies present among the SMEs (Bottenberg & Christal, 1968). The results of the hierarchical cluster analysis were used to eliminate SMEs who clustered "inappropriately." An *inappropriate clustering* was defined as a tanker or transport pilot who clustered with fighter or bomber pilots, or a fighter or bomber pilot who clustered with tanker or transport pilots.

At this point in the analysis, the SMEs for each aircraft were randomly divided into two subsamples. For each subsample of SMEs, the predicted-performance scores were averaged. Thus, each applicant profile was associated with eight composite performance measures (two subsamples × four types of aircraft). One composite performance measure for each aircraft type was entered into one of two data matrices, each with 200 rows (pilot candidates) and four columns (aircraft type or training categories). The entry in each cell of each matrix was the predicted performance of individual $i$ on job $j$.

Each of the two predicted-performance matrices was analyzed using the SAS/OR linear programming package (SAS Institute, 1989) to test for the utility of differentially assigning individuals to training categories. The objective function was to maximize mean predicted performance, with the constraints being that each individual could be assigned to only one of four jobs, and each job was constrained to a total of 50 assignments. The result of the optimization on each matrix of composite predicted-performance scores was an aircraft assignment matrix for each subsample with four columns (representing four aircraft assignments) and 200 rows (representing individuals). The entries in each of the two assignment matrices (one for each matrix of predicted-performance scores) consisted of ones and zeroes, with the ones representing job assignments. Thus, each row had only one nonzero entry (the individual's assignment) and each column had 50 nonzero entries.

Next, the assignment matrix from each subsample of SMEs was applied to the predicted-performance matrix for the other subsample. This procedure, analogous to double cross-validation in a regression analysis, was intended to minimize the effects of sampling error in estimating the effects of optimal assignment on mean predicted performance. The result of this cross-application of assignments, then, was two optimization solutions.

For each subsample, a random assignment solution was used as a baseline against which to compare the optimal assignment solutions. The random solution was chosen as a baseline because it represents a standardized although somewhat arbitrary reference point against which other more optimal solutions could be compared. In practice, the actual solution obtainable without benefit of the type of assignment information produced in this effort would probably be "better" than random assignment—or it could be worse. Because it is arguable how much better (or worse) one might do, the random solution is at least replicable and consistent with procedures for estimating effect sizes found in the general literature (e.g., Johnson & Zeidner, 1990).

# RESULTS

## Interrater Reliability

The interrater reliability analyses indicated that two of the raters were not consistent with the other SMEs of the same aircraft type. One discrepant rater was a tanker SME and the other a transport SME (see Table 2). In both cases, examination of the rater policies or regression equations indicated that each rater used only one variable, such as college grade point average, to rank candidates. Most SMEs used a number of variables in their rankings, based on the regression weights in their individual equations, which suggested that the "one-variable" SMEs may not have performed the sorting exercise as conscientiously as their peers. With data from the two discrepant SMEs removed, the interrater reliability statistics, $r_{kk}$, varied from .92 to .95 for the four aircraft types, indicating satisfactory interrater agreement.

## Intrarater Consistency

Each SME's rankings for the 200 candidates were regressed against the 21-variable predictor set. The multiple correlations for the SMEs ranged from .641 to .961, with a mean of .826, indicating a satisfactory level of within-rater consistency. Thus, no SMEs were eliminated at this stage of analysis.

## Hierarchical Cluster Analysis

A hierarchical cluster analysis indicated that the SMEs fell into one of five groups: a bomber group, a fighter group, a tanker group, a transport group, and a "generic" group. SMEs who clustered into an inappropriate aircraft

TABLE 2
Subjects Remaining at Each Stage of Analysis

| | *Aircraft Type* | | | |
|---|---|---|---|---|
| *Stage* | *Bomber* | *Fighter* | *Tanker* | *Transport* |
| Initial | 11 | 13 | 16 | 17 |
| Intrarater consistency | 11 | 13 | 15 | 16 |
| Hierarchical cluster analysis[a] | 11 | 13 | 15 | 16 |
| Final[b] | 6 | 11 | 11 | 11 |

[a]Two subjects eliminated for low interrater reliability. [b]Sixteen subjects eliminated for clustering inappropriately.

(i.e., tanker into fighter, bomber into transport) were eliminated from subsequent analyses. This procedure resulted in the elimination of 16 SMEs (see Table 2). Six of the remaining 39 SMEs were from the bomber group, and 11 SMEs from each of the other three types of aircraft were retained. Interrater reliability statistics were recomputed for the 39 SMEs remaining after the hierarchical cluster analysis. The $r_{kk}$ interrater reliability statistics were in an acceptable range (.88–.93).

## Performance Prediction Equations

Eight performance prediction equations were generated. The criterion for each regression equation was one of the eight (two subsamples × four aircraft) composite predicted-performance measures, and the predictors were the 17 scores from the data profile cards and the four binary variables computed from the preference measure. The multiple correlations for the eight equations varied from .85 to .94, indicating a high degree of relation between the mean ranking and the information on the data cards.

## Optimization

The results of the linear programming optimization analysis are shown in Tables 3 and 4, along with information from a solution that involved random assignment of candidates to aircraft type. As the data in Tables 3 and 4 indicate, optimization resulted in an overall improvement of a little more than one third of a standard deviation in predicted performance.

# DISCUSSION

The results of this study provide an indication of the degree of improvement in predicted performance that might be obtainable using an optimal assignment system for placing Air Force pilot candidates into training tracks. That

TABLE 3
Results of Optimal and Random Assignment of
Pilot Candidates to Four Training Categories (Subsample 1)

| | *Performance Indicator* | | | |
|---|---|---|---|---|
| *Aircraft* | *1. Mean Random Assignment* | *2. SD Random Assignment* | *3. Mean Optimal Assignment* | *4. Change* |
| Bomber | 25.95 | 11.32 | 26.55 | .05 |
| Fighter | 24.27 | 11.75 | 30.59 | .54 |
| Tanker | 24.87 | 12.50 | 30.94 | .49 |
| Transport | 25.08 | 10.34 | 28.21 | .30 |

*Note.* Subsample 1 optimal assignments based on solution from Subsample 2. Change = (Mean Optimal Assignment – Mean Random Assignment)/*SD* Random Assignment.

TABLE 4
Results of Optimal and Random Assignment of
Pilot Candidates to Four Training Categories (Subsample 2)

| | *Performance Indicator* | | | |
|---|---|---|---|---|
| *Aircraft* | *1. Mean Random Assignment* | *2. SD Random Assignment* | *3. Mean Optimal Assignment* | *4. Change* |
| Bomber | 24.85 | 10.76 | 27.75 | .27 |
| Fighter | 24.65 | 10.00 | 26.17 | .15 |
| Tanker | 25.50 | 11.10 | 36.72 | 1.01 |
| Transport | 25.22 | 9.73 | 26.41 | .12 |

*Note.* Subsample 2 optimal assignments based on solution from Subsample 1. Change = (Mean Optimal Assignment − Mean Random Assignment)/*SD* Random Assignment.

improvement in the available performance metric was modest, about one third of a standard deviation in performance. However, even modest increases in performance can result in substantial cost savings to an organization, as has been demonstrated in previous research (i.e., Cascio, 1991; Nord & Schmitz, 1991; Zeidner & Johnson, 1991).

For example, a gain in mean predicted performance equivalent to that observed for optimal assignment could theoretically be achieved with stricter criteria for graduation from pilot training. That is, a gain in mean performance of .38 of a standard deviation could be achieved by eliminating pilot candidates at the lower end of the expected performance distribution. Use of the Naylor–Shine tables suggest that assigning only the top 78% of pilot trainees to aircraft assignments would achieve results comparable to those gained through optimal assignment. However, to produce the same number of pilots, the Air Force would have to enter into training more candidates and eliminate an additional 22% of them. Thus, to achieve an increase of .38 of a standard deviation in performance for the approximately 600 pilots the Air Force trains in a year, an additional 169 pilots would need to be accessed into pilot training at costs currently in excess of $250,000 per graduate.

Future directions for research include replicating the results of this study using a different criterion measure. For this research, pilot candidates are being tested prior to entry into specialized training tracks. At the end of training, performance ratings are being collected. Thus, test scores and other predictor information will be evaluated against to empirical criteria as compared to the SME rankings in this study.

In addition, two other types of studies are being conducted. One study is examining the utility of classifying pilot and navigator candidates into entry-level training. Other research is addressing methodologies for examining the utility of classification procedures, insofar as most utility analyses are based on selection procedures. It is expected that, together with the present study, this program of research will result in improved methods of utilizing Air Force aviation personnel.

## ACKNOWLEDGMENTS

Portions of this article were presented at the meeting of the Society for Industrial and Organizational Psychology, April 1992, Montreal, Quebec.

We thank Joseph Weeks and David Perry, Jr., for their helpful comments. We also gratefully acknowledge the technical support provided by Bill Glasscock, Rob Long, Fernando Sanchez, Eric Miles, and Sandy Buchanan.

The views expressed in this article are those of the authors and do not necessarily reflect official views of the Air Force or the Department of Defense.

## REFERENCES

Bottenberg, R. A., & Christal, R. E. (1968). An iterative technique for clustering criteria which retains optimum predictive efficiency. *Journal of Experimental Education, 36*(4), 28–34.

Brogden, H. E. (1954). A simple proof of a personnel classification theorem. *Psychometrika, 19,* 205–208.

Carretta, T. R. (1990). *Cross-validation of experimental USAF pilot training performance models* (AFHRL–TR–89–68). Brooks Air Force Base, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Cascio, W. (1991). *Costing human resources.* Boston, MA: PWS–Kent Publishing.

Christal, R. E. (1968). Selecting a harem—and other applications of the policy-capturing model. *The Journal of Experimental Education, 36,* 35–41.

Christal, R. E., & Weismuller, J. J. (1976). *New CODAP programs for analyzing task factor information* (AFHRL–TR–76–3, AD–A026 121). Lackland Air Force Base, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory.

Dougherty, T. W., Ebert, R. J., & Callender, J. C. (1986). Policy capturing in the employment interview. *Journal of Applied Psychology, 71,* 9–15.

Flanagan, J. C. (Ed.). (1948). *The aviation psychology program in the army and air forces* (Army Air Forces Aviation Psychology Program Research Reports No. 1). Washington, DC: U.S. Government Printing Office.

Horst, P. (1956). Multiple classification by the method of least squares. *Psychometrika, 12,* 3–16.

Johnson, C. D., & Zeidner, J. (1990). *Classification utility: Measuring and improving benefits in matching persons to jobs* (IDA Paper P–2240). Alexandria, VA: Institute for Defense Analyses.

Naylor, J. C., & Wherry, R. (1965). The use of simulated stimuli and the JAN technique to capture and cluster the policies of raters. *Educational and Psychological Measurement, 25,* 969–986.

Nord, R., & Schmitz, E. (1991). Estimating performance and utility effects of alternative selection and classification policies. In J. Zeidner & C. D. Johnson (Eds.), *The economic benefits of predicting job performance: Vol. 3. The gains of alternative policies* (pp. 73–137). New York: Praeger.

SAS Institute, Inc. (1989). *SAS/OR User's Guide* (Version 6, 1st ed.). Cary, NC: SAS Institute.

Skinner, M. J., & Ree, M. J. (1987). *Air Force Officer Qualifying Tests (AFOQT): Item and factor analysis of Form O* (AFHRL–TR–86–68). Brooks Air Force Base, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Thorndike, R. L. (Ed.). (1947). *Research problems and techniques in the Army Air Forces* (Aviation Psychology Program Research Reports, No. 3). Washington, DC: U.S. Government

Printing Office.

Ward, J. H. (1958). The counseling assignment problem. *Psychometrika, 23,* 55–65.

Zeidner, J., & Johnson, C. D. (1991). *The economic benefits of predicting job performance: Vol. 3. The gains of alternative policies.* New York: Praeger.